DOCUMENT RESUME

ED 435 184 FL 026 036

AUTHOR	De Avila, Edward
TITLE	Setting Expected Gains for Non and Limited English Proficient Students. NCBE Resource Collection Series.
INSTITUTION	National Clearinghouse for Bilingual Education, Washington, DC.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	1997-11-00
NOTE	27p.
CONTRACT	T295005001
PUB TYPE	Guides - Non-Classroom (055)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Achievement Gains; *Educational Objectives; Elementary Secondary Education; *English (Second Language); *Language Proficiency; Language Tests; *Limited English Speaking; Teaching Methods; Testing

ABSTRACT

Programs designed to improve the language proficiency of limited-English-proficient (LEP) students have had mixed results. Much confusion has arisen from variable approaches to what can be expected of the programs. Recently, "expected gain" has become an important concept in documenting the educational development of LEP students, requiring analysis of the relationship between quality of instruction and measurable student outcomes. Key factors in this concept include: assessment outcomes; the setting of reasonable individual expectations and sensitivity to growth; and the effectiveness of instructional practices and programs. Psychological and pedagogical implications follow from each of these factors. A discussion of "expected gain" looks at research on the relationship between LEP student gains in language proficiency and probability of academic success in mainstream programs, the importance of how and when the student enters the program, and the role of appropriate measurement in assessing both expected and actual gain. Additional considerations include quality of instruction, differences inherent in elementary and secondary school levels, and variations in student population. Contains 18 references. (MSE)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 435 184

NCBE
RESOURCE COLLECTION SERIES

SETTING EXPECTED GAINS for Non and Limited English Proficent Students

Edward De Avila, Ph.D.

November
1997

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

ERIC

Acknowledgements

Special thanks go to Drs. Sharon Duncan, Jill Morgan and Cecilia Navarrete for their thoughtful comments and review of previous versions of this paper.

The National Clearinghouse for Bilingual Education (NCBE) is funded by the U.S. Department of Education's Office of Bilingual Education and Minority Languages Affairs (OBEMLA) and is operated under Contract No. T295005001 by The George Washington University, Graduate School of Education and Human Development, Center for the Study of Language and Education. The contents of this publication are reprinted from the NCBE Resource Collection. Materials from the Resource Collection are reprinted "as is." NCBE assumes no editorial or stylistic responsibility for these documents. The views expressed do not necessarily reflect the views or policies of The George Washington University or the U.S. Department of Education. The mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. government. Readers are free to duplicate and use these materials in keeping with accepted publication standards. NCBE requests that proper credit be given in the event of reproduction.

Director: Joel Gomez
Associate Director: Minerva Gorena
Assistant Director: Barbara Silcox

Setting Expected Gains for Non and Limited English Proficient Students

Edward De Avila, Ph.D.

November 1997

The primary instructional vehicle in the schools is language. Unfortunately, however there are large numbers of students in the United States who find it difficult to benefit from instruction because of limited proficiency in the language of the classroom, usually English. These children come from a wide variety of environments. There are several million such children and even more adults. They share the fact that they have not been a part of the mainstream linguistic environment, and, as a result, have been excluded from both the educational and social benefits. They may be economically poor and therefore linguistically isolated from the mainstream or they may come from backgrounds where the home language or dialect does not match that of the schools. In either case, or, for whatever reasons, they are unable or find it difficult to benefit from the mainstream instruction offered in the schools and do not seem to learn.

Over the past twenty years, programs designed to improve the language proficiency of limited English proficient (LEP) students have met, unfortunately, with mixed results. And, largely because of a lack of adequate evaluative documentation, results have been equivocal regardless of program quality. Much of the confusion has come out of variable approaches to the concept of growth and what can be expected from programs of this type.

Recently, "expected gain" has become an important concept in documenting the educational development of limited and non English proficient speaking students. An understanding of this concept requires an analysis of the relationship between quality of instruction and measurable student outcomes. The three key factors which both underlie this relationship and provide the necessary foundation upon which expectations for learning can be derived or generated in a meaningful and defensible manner are:

- **Assessment Requirements**

  The concept of expected gain, as implemented in programs for non and limited English proficient (LEP) students, assumes a direct empirical relationship between student gains in language proficiency and their probable success in a mainstream program. In this context, valid and reliable assessment of growth in language proficiency, sensitive to gains that are attributable to program and instruction, are essential. Tests failing to exhibit this relationship will not provide a stable basis for either setting expectations or demonstrating growth.

- **Setting Expectations and Sensitivity to Growth**
  Setting a reasonable "expectation" for student performance must be done on an individual basis beginning with a determination of where the student enters the program and measuring growth in increments sensitive to substantive linguistic changes.

- **Instructional Practices and Programs**
  The process of setting expectations for growth necessarily assumes exposure to an effective instructional program. Obviously without a quality program, expectation levels are meaningless, regardless of how they are created. On the other hand, properly set expectations coupled with effective programs can go a long way towards creating a positive educational environment as well as documenting and/or validating programs.

Psychometric and pedagogical implications follow from each factor. However, before discussion of these implications, it would seem important to first clarify what is meant by language proficiency since much of the confusion over the effectiveness and purposes of programs aimed at "remediating" limited English proficiency results from a lack of a clear definition. As will be seen, in this connection, the distinction between language proficiency and academic achievement will become critical to this and any discussion regarding language minority children.

Language Proficiency Defined

In as much as the concept of language proficiency in this context is directly related to the concept of expected-gain, it needs to be defined both conceptually and empirically. The term "language proficiency" as it is used here refers to those linguistic elements necessary for successful communication within the school environment. It is a broader concept than the concept of "academic achievement," though it underlies success in school. Thus, while language proficiency is viewed as a necessary element in defining academic success in the mainstream, it is not, in itself, sufficient to guarantee success as defined by performance which is indistinguishable from that of mainstream students.

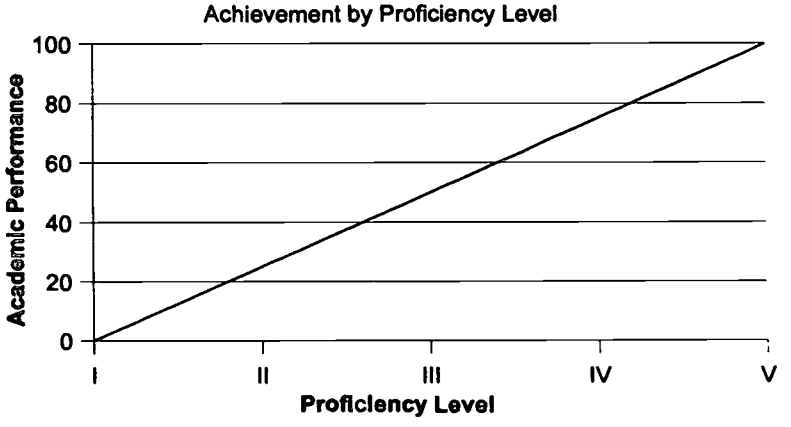
Defined as communication, language proficiency consists of both receptive and productive skills, input and output, information sent and received. It is made up of both oral and literacy skills: listening, speaking, reading and writing. Proficiency in each of the four domains is viewed as a necessary element to language proficiency, as it contributes to academic success in the specific sense. Language proficiency is a necessary element to success in the general sense but not sufficient in the specific sense of guaranteeing success in school.

Knowing that a student is linguistically proficient tells us that s/he is able to benefit from instruction in the language of the classroom. While a test of language proficiency tells us nothing about how well a student will perform on a test of American history, it will, however, tell us that s/he can understand or comprehend (listen to) oral instruction on American history. Moreover, it will tell us whether s/he can be expected to comprehend and obtain textual information (reading) on American history, as well as write and speak about what s/he has learned about American history.

Relationship between student gains in language proficiency and probability of academic success in mainstream classrooms

Language proficiency is made up of both oral and literacy skills. Let us first consider oral skills. There are several studies that apply to the present discussion. The first was conducted in 1978 under contract to the California State Department of Education. In this study De Avila, Duncan and Cervantes (1978) hypothesized a linear relationship between five levels of a widely used test of oral language proficiency and academic performance as measured by the CTBS-U (see Figure 1).

Figure 1

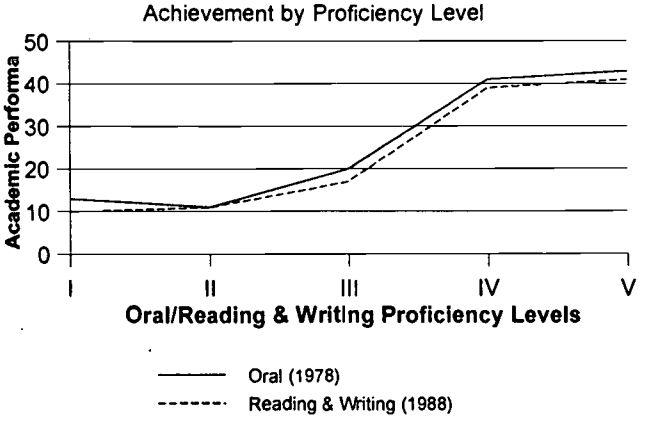


As predicted, oral language proficiency was found to be a significant predictor of academic performance. Researchers found that students scoring at Levels 4 and 5 on the oral test passed the CTBS-U at or above the 36th percentile (see Figure 2). In other words, oral proficiency was found to be a necessary element for success. The results of this study were, in part, used by several State Departments of Education to set "reclassification" or "cut-off" scores for determining student eligibility for bilingual programs.

In a further attempt to test the assumed relationship between academic performance and literacy (reading & writing), the above study was replicated in 1988 using a reading and writing test in place of the oral test used in the 1978 study. Results from both studies are shown on Figure 2. The similarity of results is striking.

The same fundamental results were obtained as in the first study. This time, however, a direct relationship was found between language proficiency, as measured by literacy, and academic performance contrast to the same relationship between oral proficiency and academic achievement found in the first study. Moreover, it was found that of students passing the reading and writing test at the "competent literate" Level (3), over 90 percent passed the CTBS-U at or above the 36th percentile (see Figure 2).

Figure 2



These studies, along with numerous other studies conducted by various researchers over the past fifteen years, provide ample support for the hypothesized relationship between language

proficiency and school performance as well as the justification or basis for examining the extent of "expected growth" over time in relation to academic performance.

It should be obvious that the choice of a valid and reliable test of language proficiency is critical. Tests failing to "predict" performance in the manner discussed here would certainly be problematic in actually setting an expectation level or score.

There are, therefore, two primary requirements that must be placed on whatever test of language proficiency is used to measure growth. First it must predict or be related to programmatic criteria, be it defined as achievement on a statewide standards or a nationally normed test. Secondly, it must produce increments of growth in units which are reflective of learning and educationally meaningful.

Where the Student Enters the Program

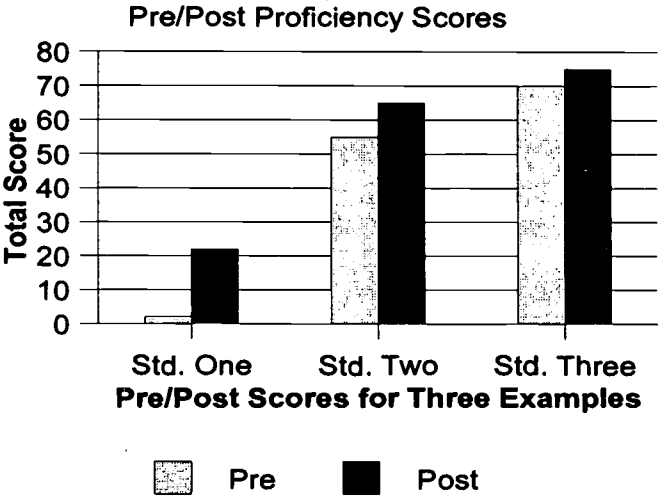
It cannot be assumed that all students will learn at the same rate or to the same extent. In large measure, extent of growth is limited by how far along the student is on the learning curve when s/he enters or begins a program.

It is of key importance to understand that there is a difference between *expected growth* and *possible growth* when setting expectations. Stated in another way, this means that if a student begins at zero on a 0-100 scale, possible growth is 100. Conversely a student who begins at 99 can only grow to 100, an improvement of only one point; in sharp contrast to the student who had the possibility of gaining 100 points.

It is exactly because possible growth is a direct function of where, along the program/measurement continuum a student begins that it is essential to establish that point before setting "expected growth." It would be foolish to expect the same growth for all students regardless of entry point, program type, quality or effectiveness. Given that not all students can be expected to show the same amount of growth in the same time frame, it becomes essential to establish the point at which the student begins. In addition, regardless of what measurement device is used, it must be sensitive enough to show growth in meaningful increments. Thus, the choice of an appropriate "metric" becomes critically important. An improper metric such as categorical or nominal scales can obscure growth. Figure 3 illustrates three examples where

student progress or growth is tracked both by level of proficiency (1 to 5) and by continuous score (0 to 100).

Figure 3



Consider Student One who begins knowing absolutely no English at all, a Level 1 student on a 1 to 5 scale. As shown on Figure 3, Student One made "no change" in proficiency level between pre-and post-tests. On the other hand, however, the total score for Student One shows a gain of 20 total points along the proficiency continuum. Student Two, in contrast, gained a full level between pre- and post-tests, however, gained only 10 points along the proficiency continuum. Similarly Student Three gained a full level, but showed only a five point increase between tests.

"Level," in the above, indicated that only two of the three examples showed gain. Examination of the total point scored revealed, perhaps ironically, that the student who showed the greatest gain (in points) made no gain or change in "level."

The above example illustrates the importance of why the choice of an appropriate metric to indicate change, gain or learning, becomes important. The use of a metric such as "level" in the above example, may be insensitive to show actual growth or change.

Given different entry or starting points, what would be an acceptable "expected gain?" The following study shows how the above ideas have been applied in a large urban context. Los Angeles Unified School District has examined the "rate of growth" within the district's bilingual and ESL programs. These "gains" can be thought of as the result of a year's program intervention. Consistent with the discussion above, it should be noted that absolute growth is to a large extent a function of initial level. Thus, one would expect greater gains for an entering student than would be expected for a student further along. Certainly, one cannot expect the same growth indefinitely; any learning curve will exhibit diminishing returns.

The data used to generate the expected gains shown on Table 1 were based largely on work conducted by Toni Marsnik at LAUSD in which data on several thousand elementary level students were examined (see Figure 4). It should be noted that the "gain" scores shown below are based on continuous total scores, which are more "sensitive," and not on Levels which are less able to show change.

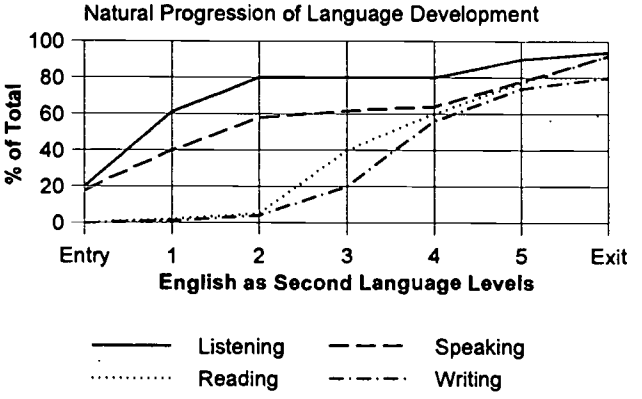
Table 1. **Average Expected Gain as A Function of Initial or Entering Level**

Level/.Lang. Proficiency	Oral	Literacy
Level 1	20	30
Level 2	10	15
Level 3	5	
Level 4		
Level 5		

Gains beyond Level 3 for the oral test are difficult to anticipate since scores at or above Level 4 are indicative of "native-like" proficiency and not as subject to program intervention or change as are scores at the lower levels of proficiency. In other words the test reaches a "ceiling effect"at this point; as it was perhaps not designed to discriminate between student's achievement levels beyond this level of proficiency. In other words, at this point language proficiency ceases to be a predictor of achievement; and it reaches a "ceiling". Therefore, low achievement beyond this level of language proficiency can no longer be associated with limited

proficiency. Similarly, while there may be slight differences in proficiency, they are not necessarily predictive of differences in achievement performance. Growth in Reading and Writing is more difficult to anticipate than changes in Oral Proficiency since changes in literacy are more directly tied to instruction and less a function of "informal" instruction. Growth in this area is therefore slower as shown in Figure 4.

Figure 4



One of the more important results of these studies is that the skills in reading, writing, listening and speaking cannot be assumed to improve at the same rate. Moreover, growth in language proficiency is largely a function of program participation and the quality of the program. Unfortunately, these data were not available. In the above studies, data were collapsed (averaged) across programs and level of participation. In effect, these data might well be described as "random" treatment effects. Certainly more detailed investigation is warranted where both student and program characteristics are examined.

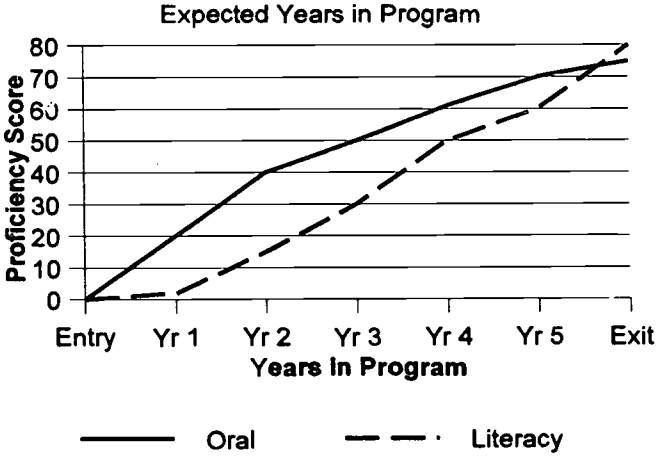
Quality of Instruction

Several inferences can be drawn from the above discussion. Perhaps the most important pedagogical implication speaks to the average time a student can be expected to require special program treatment. For example, the student who enters a program knowing no English whatsoever (See Figure 5), a student at the "entry level" of an ESL program, would be a (language proficiency index = 1/1) student who scored at chance levels on both oral and literacy tests. Consider first oral development since it precedes literacy development.

According to the values shown on Table 1, a 1/1 student would be expected to gain approximately 20 points in oral development in the first year, from 15 (approximately chance) to 35 points for example, still Level 1 (Level 1 = 0 to 54). In the second year, the student would gain another 20 points for a total of 55, a low Level 2. In the third year, the student would move from a low level 2 to a high level 2, or a total of about 65 points. In the fourth year the student could be expected to move into Level 3 with a total of approximately 70 points. Finally, full native-like oral proficiency would not be "expected" until completion of the fifth year.

Literacy development would follow the same basic pattern or steps. Our research (cited above), however, shows that the development of literacy skills is somewhat slower at the lower levels. However, once minimal oral skills have been established, students move quickly through the middle levels as shown by the slope of the curves in Figure 5.

Figure 5



It is noteworthy that the sum of the expected gain values across the different levels is 35 (20+10+5) for oral and 45 (30+15) for the reading and writing which averages out to approximately 13 total points. Thirteen points translates into approximately one proficiency level per year. The critical point, however, is that while it may seem reasonable to set expectations on the basis of a level per year, it would be misleading as discussed above.

Finally, it is also worth bearing in mind that an approach based on differential expectations can offer a powerful metric for evaluating both student progress and programmatic effectiveness. Student growth has been discussed. Programmatic effectiveness in this context becomes a

matter of counting the numbers or percentage of students obtaining or exceeding their individual goals. In the ideal program, all students would reach their "expected gain."

Implications and Cautions

There are several important points to be taken from the above exercise. The first is to recognize that what has been said applies to any language and any set of tests. The relationships described above are the same for all languages. In terms of program evaluation, the approach works equally well in a dual language program as well as in a program directed toward the improvement of one language. Similarly, any test should be held to the same standards of validity, reliability and ability to make distinctions as described above. Tests should provide predictive information while, at the same time, measure change in meaningful increments. Tests that do not meet these two conditions should not be used to set expectations nor measure growth.

As was seen in Figure 4, there is very little growth in literacy skills in the first two levels of ESL study. On the other hand, growth in oral skill is rapid, particularly, listening skills. It would, therefore, appear that the acquisition of English as a second language develops in a non-linear fashion. Therefore, according to these data, initial programmatic emphasis, at least at the elementary level, should be directed toward the development of beginning oral skills before seriously undertaking reading and writing.

In this connection it is critical to note that the dat cited above were all based on elementary level students and we would not expect the same values to hold at the secondary level. In fact, based on work conducted in 1988, there is reason to believe that proficiency develops somewhat differently at the secondary level than it does at the elementary level. For example it has been found that elementary level students who are unable to speak a language (English and Spanish) are seldom (almost never) able to read and write in that language. On the other hand, there are significant numbers of students at the secondary level who are able to read and write in a second language while not being able to converse in the language. These students tend (1) to be recent arrivals as opposed to second and third generation students; (2)to be educated in the home language; and (3) have received instruction in the second language in the homeland. The predictions for these students would be very different than for others.

It is also important to note that, in the final analysis, full scale language proficiency requires proficiency in all four of the linguistic domains discussed. Since language proficiency, as it has been used here includes both speaking and writing skills, it would not be surprising to find there are a good many students from English-only backgrounds who are of limited English proficiency.

Note that the above expected gain values are based on averages that cannot be used to form conclusions or mark the progress of an individual student. Stated in another way, group data cannot be proved or disproved by a single example; results must be evaluated on an average basis. The implicit model underlying the approach taken here is based on probabilities and are accurate only to the extent of the probability value. To say that the prediction is accurate nine out of ten times is no different than to say that the prediction is inaccurate one out of ten times. Thus, the approach as well as any program predicated on the approach, must be evaluated based on the total group performance and cannot be proven or disproven on the basis of a single example or individual's test scores.

As a further caution, it should also be borne in mind that even though two tests may employ or report scores using the same metric, there may be distinct differences depending on the norming sample. Thus, the 40th percentile on one test may not have the same meaning as the 40th percentile on another test. Without a common reference group, such comparisons are, at best, difficult, and, possibly misleading at worst.

On a practical side, the determination of program eligibility or placement based on percentile values taken from a test different from the one used by a local district is problematic, particularly if one or the other has not been validated in terms of an external criteria.

Thus, it is quite possible that a child might be exited from a program before s/he is ready or denied access to needed programs when they are in need. For example, the State Department of Education Guidelines may mandate that students performing above the 30th percentile on a test of language proficiency are either ineligible or no longer eligible for services. The assumption is that the 30th percentile on the language proficiency test is somehow equivalent or predictive of a score on an achievement test. While it may be the case that the particular test used to set the 30th percentile as the "cut-score" corresponds to the desired levels of performance on the criterion test, there is no guarantee that the same relationship would hold

for the proficiency test used locally. In other words, without equating the scales from different tests there is a strong probability that educational decisions would be based on misaligned rulers.

While the preceding discussion on the various relationships between proficiency and academic performance-over-time has been suggestive, it is not definitive. What is needed to ensure that the above expectations are reasonable would be a series of longitudinal studies across age, time in program, program type, and measurement instrumentation.

The studies cited in the above discussion have not included any information on either program differences or methods of measurement. Thus, it would be difficult to conclude that the same cut scores or expectations would hold across all variations in programs and measurement techniques. A possible approach to the study of the problem would be to locate districts with sufficient longitudinal data on which to conduct a series of post-hoc analyses where "growth" is plotted over time. A major problem, of course, would be the extent to which the data are available in forms and formats which are amenable to analyses. Moreover, there would be no control over the quality of the data supplied by individual districts. Although, while not ideal, a post hoc approach would certainly be an improvement over the current situation.

Finally these results coupled with those of future studies would help in resolving recent debate over how long it takes to become proficient in English. The present data seem to suggest that it takes approximately five to seven years. However, until further work is completed, the issue is subject to continuing debate.

References

De Avila, E., Cervantes, R. & Duncan, S. (1978) *Bilingual Exit Criteria.* CABE Research Journal, 1(2).

Language Assessment Scales. (1988) *Oral Technical Report.* CTB McGraw/Hill Monterey, CA.

Additional Resources

Anstrom, K. (1997). *Academic achievement for language minority students: Standards, measures, and promising practices*. Washington, DC.: National Clearinghouse for Bilingual Education. [www.ncbe.gwu.edu/ncbepubs/acadach.html]

Council of Chief State School Officers Stanford Working Group (1996). *Systemic Reform and Limited English Proficient Students.* Washington, DC.: Author [www.ccsso.org/pdfs/srandlep.pdf]

Del Vecchio, A. et al. (1994). *Whole-school bilingual education programs: Approaches for sound assessment.* Washington, DC.: National Clearinghouse for Bilingual Education. [www.ncbe.gwu.edu/ncbepubs/pigs/pig8.html]

Geisinger, K.F. (1992). Testing limited English proficient students for minimum competency and high school graduation. *Proceedings of the second national research symposium on LEP student issues: Focus on evaluation and measurement. Volume 2.* Washington, DC.: United States Department of Education, Office of Bilingual Education and Minority Languages Affairs.

Hernandez, R.D. (Winter, 1994). Reducing bias in the assessment of culturally and linguistically diverse populations. *The Journal of Educational Issues of Language Minority Students,* 14, pp. 269-300. [www.ncbe.gwu.edu/miscpubs/jeilms/vol14/hernand.html]

Hopstock, P. J. (1995). *Recommendations on student outcome variables for limited English proficient (LEP) students.* Arlington, VA.: Development Associates, Special Issues Analysis Center. [www.ncbe.gwu.edu/miscpubs/siac/outcome/index.html]

McLaughlin, B. et al. (1995). *Assessing language development in bilingual preschool children.* National Clearinghouse for Bilingual Education: Washington, DC. [www.ncbe.gwu.edu/ncbepubs/pigs/pig22.html]

National Clearinghouse for Bilingual Education. (1997). *High stakes assessment: A research agenda for English language learners. Symposium summary.* Washington, DC: author. [www.ncbe.gwu.edu/ncbepubs/report/highstakes.html]

Navarrete, C. and Gustke, C. (1996). *A guide to performance assessment for linguistically diverse students.* Albuquerque, NM.: EAC West, New Mexico Highlands University. [www.ncbe.gwu.edu/miscpubs/eacwest/perform.html]

Olson, J.F. and Goldstein, A.A. (1997). *The inclusion of students with disabilities and limited English proficient students in large-scale assessments: A summary of recent progress.*

Washington, DC.: National Center for Education Statistics, U.S. Department of Education.. [www.ed.gov/NCES/pubs97/97482.html]

Pierce, L.V. and O'Malley, J.M. (1992). *Performance and portfolio assessment for language minority students.* Washington, DC.: National Clearinghouse for Bilingual Education. [www.ncbe.gwu.edu/ncbepubs/pigs/pig9.html]

Saville-Troike, M. (1991). *Teaching and testing for academic achievement: The role of language development.* Washington, DC.: National Clearinghouse for Bilingual Education. [www.ncbe.gwu.edu/ncbepubs/focus/focus4.html]

Short, D. (1993).Assessing integrated language and content instruction. *Tesol Quarterly*, Vol. 27, No. 4. [www.ncbe.gwu.edu/miscpubs/tesol/tesolquarterly/assessin.html]

Switzer, F. L., et al. (1994). *Assessment of English language learners in Chapter I Programs.* Arlington, VA.: EAC East, George Washington University. [www.ncbe.gwu.edu/miscpubs/lists/chapter1.html]

Zehler, A. (1994). *An examination of assessment of limited English proficient students.* Arlington, VA.: Development Associates, Special Issues Analysis Center. [www.ncbe.gwu.edu/miscpubs/siac/lepasses.html]

NCBE *Newsline*

To get the latest information relating to funding opportunities, legislation, current research, and upcoming meetings, subscribe to NCBE's electronic newsletter, *Newsline*. To subscribe, send an e-mail message to: majordomo@cis.ncbe.gwu.edu. In the body of the message type: **subscribe newsline**

EPILOGUE

The following comments represent the author's thoughts related to:

Setting Expected Gains for Non and Limited English Proficient Students

How Long Does It Really Take to Learn English?

Ed De Avila, Ph.D.

Students from non-English-speaking backgrounds are normally classified as non-limited and proficient speakers of English according to their level of English proficiency. An important question for schools becomes how many students move from one level to the next and how long does it take?

In a recent brief article De Avila (1997) hypothesized an inverse linear relationship between expected growth in English language proficiency and initial proficiency. That is, that the amount of expected gain between two test administrators was to a large extent a function of initial proficiency; the greater the initial proficiency, the less the expected growth. According to the position, the use of categories or levels of proficiency would be increasingly insensitive to the growth as proficiency increased. It was further argued that units of change or analysis had to be based on equal interval scales made up of units or scores sufficiently sensitive to detect small as well as gross changes in proficiency. Therefore, it was argued, that the common practice of expecting growth of one level per year is perhaps unreasonable and tends to obscure actual growth. Limited data were presented in support of the hypothesis. However, the data in the earlier study were restricted to categorical or proficiency levels only. While observing levels only tended to limit the results, a number of important implications were suggested concerning the educational treatment of children from non and limited English speaking backgrounds.

The purpose of this current small study was twofold. The first goal was to examine the relationship between "time" and "oral language proficiency," assessed by means of the Pre-LAS, a commonly used test of oral language proficiency. The second purpose was to examine "expected gain" across "time" as a function of initial starting point as hypothesized by De Avila (1997).

In this earlier study De Avila worked backwards from pre-post data collected by Toni Marsnik and her colleagues in Los Angeles over the past few years. Using these data which were limited to proficiency level categories, it was hypothesized that students beginning at LAS Level 1 would gain approximately 20 raw score points, students at Level 2 would gain approximately ten raw score points, and students at Level 3 would gain about five raw score points. While the hypothesis was supported in the sense of establishing the linear relationship between expected growth and time, the score predictions were little more than educated guesses.

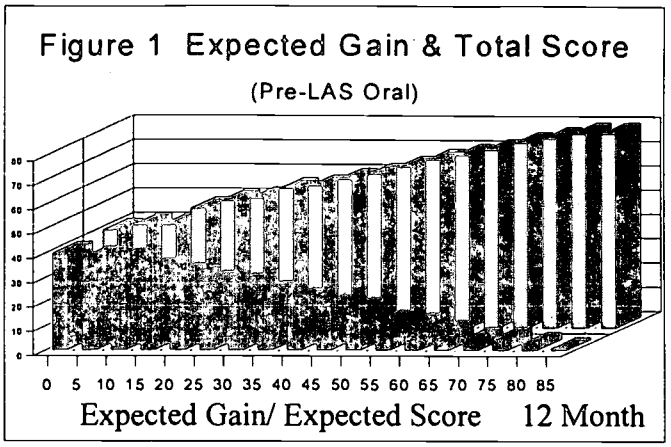
The data collected in the present study included total scores as well as proficiency level categories, enabling a far more precise estimate as to "expected gain." Thus, the data reported below represent a small-scale longitudinal sample.

Method: A total of 203 children between the ages of 54 and 80 months were administered two versions of the Pre-LAS test of oral language development. The time interval for the two test administrations was between three and sixteen months. Data on the first test were collected from school records. Data on the second test were collected in the normalization of a new parallel version of the Pre-LAS test (Pre-LAS 2000).

The development of the statistical procedures used to generate the "expected gains" followed two steps. In the first, a score was produced by calculating the difference between the two tests administered (T2-T1). This produced a "difference score."

In the second step, the "difference score" was then "regressed" against first-test standardized total scores. Scores generated from the resultant equations were then plotted against total LAS scores as shown in Figure 1. The horizontal axis shown on Figure 1 represents the "expected" LAS Total Scores. "Expected" total scores were calculated by adding "expected gains" to initial scores.

Given the dimensions of "time" (indicated by two separate test administrations) and "initial score" (indicated by scores on the horizontal axis) the following holds: a student with an initial score on the Pre-LAS of 5 can, on average, be "expected" to gain somewhere around



40 raw score points. Adding the expected gain to the initial score results in an expected score of 45. A student with an initial score of 65 standard score points would be expected to gain approximately seven standard score points for a total of 72. A student with an initial score of 75 would be expected to gain only about 5 standard score points and have an expected score of 80 points.

Note that possible LAS total scores shown on Figure 1 indicate a range from zero to 85 points. The actual test range is from zero to 100 points. The fact that scores in the present context range between zero and 85 would indicate that the model holds only to about 80 points. Beyond this level, first test results cease to predict "gains." This is no doubt due to the inherent ceiling effect in the tests; 80 percent is basic proficiency.

It would be interesting to speculate on how long it would take to become a proficient speaker if one were to start at zero. According to these data, assuming a student began with zero proficiency, one would expect a gain of 40 points in the first year, 20 points in the second year (total = 60) and 13 points in the third (total = 73) which would leave the student just below the "proficient" speaker category cut-off score (77 at age 4, 82 at ages 5-6).

While all of these findings are encouraging as well as consistent with the original model described in the earlier paper, a significant number of cautions must be borne in mind. A further breakdown of these data are instructive. Of particular interest is the effect of the length of time or interval between test administrations. This issue can be addressed from several points. The first concerns the psychometric issue of test-retest reliability. The second, which is of greater importance in the present context, deals with the general accuracy of the model across time.

It is important to bear in mind that gain in this study would be difficult to determine since the test interval was short for some (more test-retest than longitudinal) and long for others. One way to test the extent of consistency across time would be to examine the correlation between the two test administrations as well as the residuals indicated in the regression analyses. Thus, for example, the overall correlation between the two test administrations was .82 and .83 respectively for Forms C and D.

When broken down into three test intervals, 12 months or less, 13 to 18 months, and 19 months or more, a somewhat more detailed picture emerges.

Correlations between test administrations:

Total Sample (Form C/D)	0 to 12 Months (Form C/D)	13 to 18 Months (Form C/D)	19 Months or More (Form C/D)
.76/.77	.95/.95	.70/.67	.57/.60
N= 203	68	93	36

There were six cases in which the testing interval exceeded reasonable limits. These six were dropped from the analyses.

Though perhaps not unexpected, it is noteworthy that the correlations between test administrations decreases as the interval increases. The scores will vary more the longer the student has been in a program of instruction in English. Correlations for the total sample were .76 and .77 for the two forms of the second test. Both forms were given. More interesting, the correlations for the first group (0 to 12 months) was .95 for both forms. The interesting point here is that this correlation is almost as high as the correlation between the two forms of the second test (i.e., Forms C/D, r= .98). Finally, the importance of these findings is not in that they were unexpected but that they move the field toward being able to establish empirically based expectations across a number of critical dimensions.

Given the progression indicated by the above, it could be inferred that pre-school children who begin school with virtually no English (NEP), would take about three years to master

sufficient oral skills to be virtually indistinguishable from his or her mainstream counterparts. However, while the student may have mastered sufficient oral skills to fully participate in an English speaking environment, there is no guarantee that he or she has mastered literacy or other academically related skills.

In an attempt to determine the extent of movement between test levels or placement categories across two test administrations the following analyses was conducted. Results summarized below were limited to students for whom the testing interval was between 9 and 12 months, much like the school year.

Data were available for a total of 92 students. Of these, 18 moved from *non* to *limited* categories, 9 moved from *non* to *proficient*, 38 remained unchanged and 5 showed a loss from *limited* to *non* proficient categories. Finally, of the 28 students initially identified as *proficient*, 7 moved from *proficient* to *limited*. In summary, 27 of the 38 (71%) students initially identified as either *non* or *limited* gained at least one level on the five point scale used to categorize them.

Perhaps one of the more important findings illustrated above concerns the students who showed a loss in proficiency. There are several points that can be made here. First, the losses described above may well illustrate the original point regarding the use of categorical levels in contrast to continuos scores. It may well have been that the students who showed losses in proficiency level or category had initially continuous scores that were very close to the cut-off scores, within the grey area above or below the cut-off score created by the standard error of measure. As argued above, a very small difference in scores between the initial and second test can lead to significant changes in level identification which, in turn, can be very misleading. It is entirely possible that level identification in this context could be affected by simple "regression towards the mean." Second, we know virtually nothing about the treatment of the students in the sample or the extent to which their exposure to English was constant throughout the testing interval. For example, children often migrate between the U.S. and Mexico or Puerto Rico during the course of a school year which would certainly have an impact on their acquisition of English. Without further detailed study it would be impossible to fully explain these results.

This study was limited to examining gains independent of the nature of exposure to English during the interval between the two test administrations. This information was not available. Similarly, little is known about the changes in oral proficiency attributed to maturational effects independent of second language differences. Nevertheless, results suggest that program evaluations must be based on time in the program, initial proficiency and a combination of student and program characteristics. Though not addressed specifically here, the relationship between program charactistics and development of English language proficiency is critical. Certainly, further study is needed on bilingual and ESL program s characteristics in the same way as further study on the longitudinal relationship between oral and literacy development is needed for all children.

It is also unfortunate that the current study was limited to one age group and LAS test level. It would be critical to further examine the above relationships across age. Previous studies by the author and his colleagues have demonstrated clear age differences in language proficiency and literacy (See De Avila & Duncan, 1988).

The model underlying the present data reflects the diminishing returns found in any learning curve, where initial rapid learning ultimately gives way to slower learning. It is in this sense that the approach mirrors or is analogous to normal development. It also shows how the level of effort needed to move from one point or level to the next may be greater at the upper end of the learning curve. This point may seem counter intuitive in that it implies greater expense in moving from "limited" to "proficient" categories than from "non" to "limited." The financial implications here would seem all too obvious.

Finally, the present approach is but one way to examine some of the important relationships in designing, implementing and evaluating programs for limited and non English speaking students. Since the model presented above is unabashedly empirical, it is subject to empirical test and refinement; only further research will determine its utility.

About the Author

Dr. Ed De Avila received his BA in psychology from the University of California at Berkeley. He received a Masters Degreee in Clinical Psychology from the University of Colorado at Boulder and a Ph.D. in Developmental Psychology at York University in Toronto (Ontario) Canada. Since 1976 he has been President of the Linguametrics Group, an educational research and development organization specializing in assessment.

De Avila has lectured throughout the United States and abroad. He has taught courses, served as professor and conducted research at such institutions as the University of California, University of Colorado, Stanford University, and Columbia University among others. Additionally, he has sat on a number of editorial boards, including the National Association of Bilingual Education Journal, Journal of International Psychology, American Journal of Mathematics Instruction, Journal of Applied Developmental Psychology, American Journal of Mental Deficiency, and others.

He has published numerous articles, several books, films and other educational materials, such as the Language Assessment Scales which are probably the most popularly used language proficiency tests in the country.

Over the past twenty years, De Avila has served as a consultant to a number of federal agencies both in the U.S. and abroad, as well as to numerous state departments of education, school districts, foundations, and private companies.

NCBE

National Clearinghouse for Bilingual Education
1118 22nd Street NW ▲ Washington DC 20037